

UC Davis

UC Davis Previously Published Works

Title

Systemic Metabolomic Changes in Blood Samples of Lung Cancer Patients Identified by Gas Chromatography Time-of-Flight Mass Spectrometry.

Permalink

<https://escholarship.org/uc/item/99n5r1m2>

Journal

Metabolites, 5(2)

ISSN

2218-1989

Authors

Miyamoto, Suzanne
Taylor, Sandra L
Barupal, Dinesh K
et al.

Publication Date

2015-04-01

DOI

10.3390/metabo5020192

Peer reviewed

Article

Systemic Metabolomic Changes in Blood Samples of Lung Cancer Patients Identified by Gas Chromatography Time-of-Flight Mass Spectrometry

Suzanne Miyamoto ^{1,*}, Sandra L. Taylor ², Dinesh K. Barupal ^{3,†}, Ayumu Taguchi ⁴, Gert Wohlgemuth ³, William R. Wikoff ³, Ken Y. Yoneda ⁵, David R. Gandara ¹, Samir M. Hanash ⁴, Kyoungmi Kim ² and Oliver Fiehn ^{3,6}

¹ Division of Hematology/Oncology, UC Davis Cancer Center, 4501 X Street, Room 3016, Sacramento, CA 95817, USA; E-Mail: drgandara@ucdavis.edu

² Division of Biostatistics, Department of Public Health Sciences, UC Davis School of Medicine, One Shields Avenue, Med Sci 1-C, Davis, CA 95616, USA; E-Mails: sltaylor@ucdavis.edu (S.L.T.); kmkim@ucdavis.edu (K.K.)

³ Genome Center, UC Davis, 451 Health Sciences Drive, University of California, Davis, CA 95616, USA; E-Mails: wohlgemuth@ucdavis.edu (D.K.B.); wrwikoff@ucdavis.edu (G.W.); ofiehn@ucdavis.edu (O.F.)

⁴ The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA; E-Mails: ATaguchi@mdanderson.org (A.T.); SHanash@mdanderson.org (S.M.H.)

⁵ Division of Pulmonary Medicine, Department of Internal Medicine, UC Davis Medical Center, Sacramento, CA 95817, USA; E-Mail: kyyoneda@ucdavis.edu

⁶ Biochemistry Department, King Abdulaziz University, Jeddah 21589, Saudi Arabia

[†] Current affiliation: International Agency for Research on Cancer, 69372 Lyon, France; E-Mail: barupal@gmail.com.

* Author to whom correspondence should be addressed; E-Mail: smiyamoto@ucdavis.edu; Tel.: +1-916-734-3769; Fax: +1-916-734-5356.

Academic Editor: Pollen K.F. Yeung

Received: 19 December 2014 / Accepted: 1 April 2015 / Published: 9 April 2015

Abstract: Lung cancer is a leading cause of cancer deaths worldwide. Metabolic alterations in tumor cells coupled with systemic indicators of the host response to tumor development have the potential to yield blood profiles with clinical utility for diagnosis and monitoring of treatment. We report results from two separate studies using gas chromatography time-of-flight

mass spectrometry (GC-TOF MS) to profile metabolites in human blood samples that significantly differ from non-small cell lung cancer (NSCLC) adenocarcinoma and other lung cancer cases. Metabolomic analysis of blood samples from the two studies yielded a total of 437 metabolites, of which 148 were identified as known compounds and 289 identified as unknown compounds. Differential analysis identified 15 known metabolites in one study and 18 in a second study that were statistically different (p -values <0.05). Levels of maltose, palmitic acid, glycerol, ethanolamine, glutamic acid, and lactic acid were increased in cancer samples while amino acids tryptophan, lysine and histidine decreased. Many of the metabolites were found to be significantly different in both studies, suggesting that metabolomics appears to be robust enough to find systemic changes from lung cancer, thus showing the potential of this type of analysis for lung cancer detection.

Keywords: metabolomics; lung cancer; mass spectrometry; blood

1. Introduction

Lung cancer continues to be the leading cause of cancer death for men and women in the United States in spite of reduced rates in smoking [1]. Furthermore, there is an increased global burden of lung cancer largely due to increased smoking in economically developing countries [2]. Only 16% of cases with non-small cell lung cancer (NSCLC) survive longer than 5 years, largely due to late stage diagnosis and metastasis of this disease. Early diagnosis significantly improves the 5 year survival rate for lung cancer [3]. Currently, there are no FDA-approved diagnostic tests available to detect the presence of lung cancer, especially in the high-risk smoking population. Results from the National Lung Screening Trial (NLST) showed that CT screening could reduce mortality by 6.7% [4]. However, CT screening is not cost-effective, has a high false positive rate and can expose the patient to low amounts of radiation [5]. There is also variability in the reading and interpretation of radiographic scans, thereby reducing the enthusiasm of CT screening for routine clinical use [6]. Ideally what is needed is a non-invasive blood analysis for biomarkers capable of assisting with diagnosis that might also help reduce the high false positive rate of CT scan screening [7,8].

Metabolomic changes in cancer are well-documented with increased glycolysis and decreased oxidative phosphorylation (the “Warburg effect”), as described by Deberardinis *et al.* [9]. Of particular interest are components of the glycolytic pathway, nucleotide, amino acid and fatty acid synthesis and how tumor cells are able to scavenge the available cellular and environment material to produce the necessary cellular components to support increased growth and proliferation. Metabolomic analysis has been used to distinguish between benign prostatic hyperplasia (BPH) and prostate cancer in urine samples [10], and invasive ovarian carcinoma tumors compared with borderline tumors in tissues [11], thus yielding potential biomarker panels for breast, ovarian and gastric cancers [12–16]. Maeda *et al.* used metabolomics to examine levels of plasma amino acids in the blood of patients with lung cancer [17]. Analysis of metabolic pathways has identified glycolytic and signaling pathways differentially regulated in pre-diagnostic blood samples from subjects diagnosed with breast cancer [18] and AMPK-related alterations in ovarian cancer [19]. Thus, measuring changes in metabolites detectable in

blood samples that may represent tumorigenesis has the potential to yield suspicious systemic metabolic changes indicating the presence of lung tumors [20–22]. The purpose of this study is to perform metabolomic analysis of blood samples from two separate case-control studies obtained from two different sites to determine if untargeted metabolic profiling by GC-TOF MS can identify metabolic differences in patients with lung cancer when compared with blood samples from those without cancer.

2. Results and Discussion

2.1. Differential Analysis Results of Cancer Cases and Controls

Gas chromatography (GC) time of flight (TOF) mass spectrometry (MS) was used to analyze pre-existing blood samples provided by two separate sites in pilot lung cancer case control studies. For Study 1, samples were acquired from the Fred Hutchinson Cancer Research Center (FHCRC) (Table 1A) comparing blood samples from NSCLC adenocarcinomas with controls (all current or former smokers frequency matched for age and gender). All patient samples for Study 1 (cases) were collected during a clinic visit prior to surgery for resectable early stage lung cancer and the controls were collected from clinic subjects without lung cancer. For Study 2, samples were acquired from University of California at Davis Medical Center (UCDMC) (Table 1B) and included a variety of lung cancers. For Study 1, 20 control subjects were compared to 18 cases. Data from two samples were not included in the analysis due to low analytical results for these samples.

Data results were differentially analyzed (see methods for details). In Study 1, based on p -values ($p < 0.05$), 19 metabolites differed significantly by cancer status (15 known metabolites and 4 unknown, shown highlighted in grey in Supplemental Table S1, left side). Data for 9 of the top 15 known metabolites from Study 1 with mean values, fold changes and p -values are shown in Table 2A. These metabolites are maltose, ethanolamine, glycerol, palmitic acid, lactic acid, tryptophan, lysine, histidine and glutamic acid. (Supplemental Table S2 show mean values, fold changes and p -values for all known and unknown compounds measured in Study 1).

In Study 2, 82 metabolites differed significantly for cancer *versus* control samples based on p -values ($p < 0.05$) (18 known metabolites and 64 unknown). Data for all of the metabolites (known and unknown) in Study 2 are shown in Supplemental Table S1 (middle section) with the 82 metabolites with p -value < 0.05 highlighted in grey. Results (mean values, fold change, p -values) of the same 9 metabolites as shown for Study 1 (Table 2A) are listed in Table 2B for comparison. These results show similar changes in these metabolites (increase or decrease) in both studies with the exception of glutamic acid, which shows an increase in Study 1 and a decrease in Study 2. Supplemental Table S2 show mean values, fold changes and p -values for all known and unknown compounds measured in Study 2.

Table 1. Summary of patient information for FHCRC and UCDMC samples. SE is standard error of the mean.

Type and Stage		Female	Male	Total	Smoking History
(A) Study 1 (38 Samples)					
Lung cancer	NSCLC adenocarcinoma stage unknown	8	10	18	Current or former smokers
Control		12	8	20	Current or former smokers
Average age (cancer)		62 (SE 2.38) (range 53–72)	67 (SE 3.66) (range 50–85)		
Average age (control)		64 (SE 2.71) (range 49–80)	66 (SE 2.65) (range 58–82)		
(B) Study 2 (22 Samples)					
Lung cancer		7	4	11	
	NSCLC Stage I-IIIB	1		1	1 former smoker
	NSCLC Stage IIIA-IV	2	2	4	1 never smoker 1 former smoker
	SCLC	3		3	1 unknown, 1 former smoker 1 current smoker
	Mesothelioma		1	1	1 former smoker
	Secondary 2nd metastasis to lung		1	1	1 former smoker
	other	1		1	1 former smoker
Control		6	5	11	unknown
Average age (cancer)		67 (SE 4.2) (range 47–76)	67 (SE 2.66) (range 61–73)	11	
Average age (control)		54 (SE 2.64) (range 44–61)	69 (SE 3.79) (range 61–83)	11	

Table 2. (A) Means and fold change for nine known metabolites from Study 1 that differ significantly (raw p -value < 0.05) between cancer patients and disease free controls; (B) means and fold change for nine known metabolites from Study 2.

Metabolite	Mean Cancer	Mean Control	Fold (Cancer/Control)	Raw p -Value
(A) Study 1 (FHCRC)				
Maltose	1298	780	1.664	0.013
Ethanolamine	156,214	123,699	1.263	0.016
Glycerol	66,463	49,062	1.355	0.023
Palmitic acid	53,763	41,293	1.302	0.047
Lactic acid	380,753	301,909	1.261	0.055
Tryptophan	121,513	143,383	0.847	0.005
Lysine	159,156	179,325	0.888	0.042
Histidine	30,526	37,025	0.824	0.036
Glutamicacid	39,179	27,794	1.410	0.026
(B) Study 2 (UC Davis)				
Maltose	1061	989	1.074	0.819
Ethanolamine	150,655	127,546	1.181	0.006
Glycerol	67,557	47,052	1.436	0.315
Palmitic acid	50,740	43,659	1.162	0.797
Lactic acid	381,850	296,663	1.287	0.108
Tryptophan	126,621	139,426	0.908	0.391
Lysine	167,528	172,015	0.974	0.636
Histidine	31,053	36,840	0.843	0.047
Glutamicacid	31,486	34,887	0.903	0.914

Box-plots comparing the mean intensities of 9 of the top candidate metabolites for all samples listed in Table 2A,B from Study 1 and Study 2 are shown in Figure 1. Also shown in Figure 1 are additional box plots to show a comparison between results from males (blue) and females (red) for the two studies. These plots show the same trends (increased or decreased) as all samples with the exception of glutamic acid, which shows the same trend for females, but not for males. This striking similarity between the studies is notable, especially since both studies were quite small and involved differences in the matrix (plasma and serum), dealt with different types of lung cancers and originated from different clinics. Overall, we concluded that it is possible to compare the two studies and still find evidence of an increase in glycolysis and lipid biosynthesis in blood samples, along with a general decrease in aromatic amino acids.

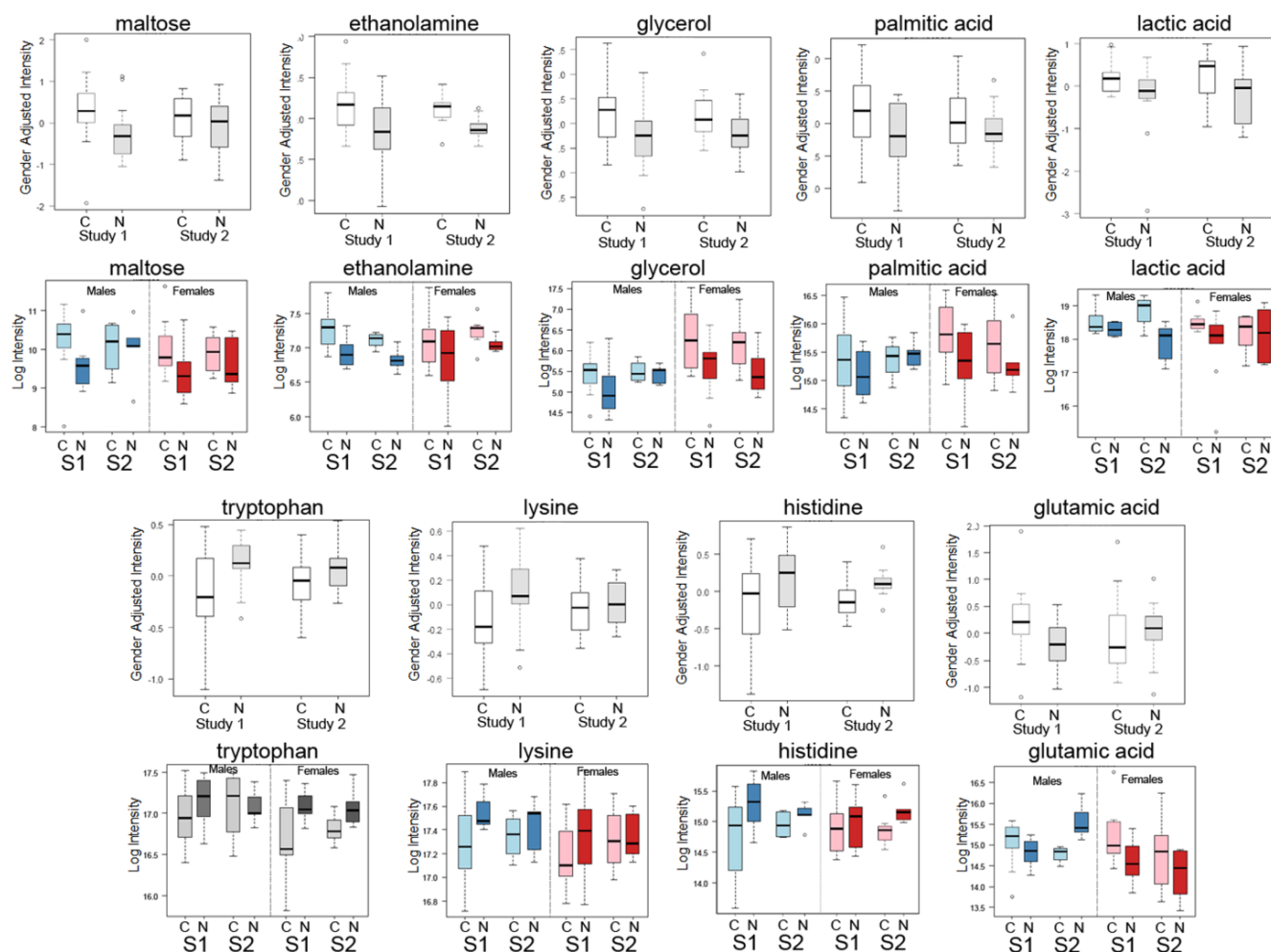


Figure 1. Box-whisker plots of top metabolite candidates in Study 1 and Study 2 with additional plots of the same metabolites from each study separated by gender (males and females). Box-whisker plots of gender adjusted intensities of top known metabolite candidates from Study 1 (S1) compared with the same compounds in Study 2 (S2) for cancer cases (C) and normal/control (N) showing similarity in the changes in both studies are shown for nine of the top metabolites: maltose, ethanolamine, glycerol, palmitic acid, lactic acid, tryptophan, lysine, histidine and glutamic acid. Shown below each metabolite plot are results for the same metabolites separated by gender. Blue plots denotes male results only and red plots denote females results only for each study.

2.2. Multivariate Analysis of Data by PLS-LDA

Multivariate analysis [23] using partial least square (PLS) [24] with linear discriminant analysis (LDA) with and without adjusting for age and gender was also performed to determine whether the blood metabolome as a combination of all metabolites identified in this study could discriminate cancer cases from control samples. PLS was used to reduce the 437 spectral peaks, each representing a metabolite, to a smaller number of latent components that distinguished cancer cases and controls and then determined which peaks were most influential in separating cases and controls as possible biomarkers. Metabolomic results from the cases and controls of Study 1 were separated by the first and second components when

adjusted for age and sex (Figure 2A) that also showed good separation when the results were not adjusted for co-variants (Figure 2C). Based on leave-one-out cross-validation (LOOCV), we attempted to assess the performance of the metabolites to correctly identify the cancer cases. Sixty-three % of Study 1 samples were correctly classified, with 66.7% sensitivity and 60% specificity when two latent components were used to predict cancer status with adjustment for age and gender (Table S3A) and without adjustment for age and gender (Table S3B).

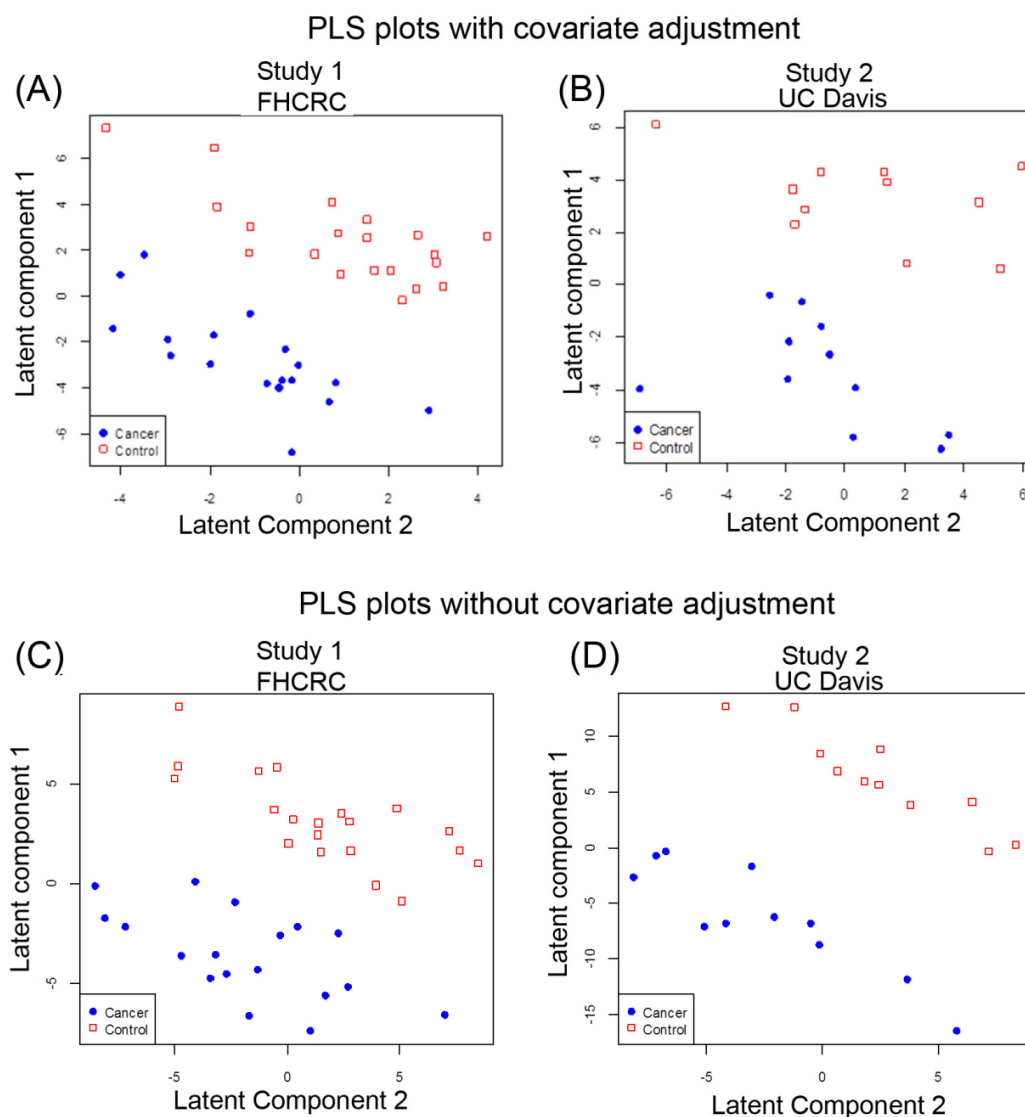


Figure 2. Multivariate PLS separates lung cancer patients and controls in two independent studies by the global metabolomic profiles. (A) PLS of Study 1 data results with gender and age adjusted; (B) PLS of Study 1 without gender and age adjusted; (C) PLS of Study 2 with gender and age adjusted; (D) PLS of Study 2 without gender and age adjusted. Red squares denote control cases and solid blue circles denote cancer cases.

For Study 2, PLS-LDA analysis also showed separation for cancer based on metabolic profiles with adjustment for gender and age (Figure 2B) as well as for metabolic profiles without adjustment for gender and age (Figure 2D). Based on LOOCV, 68% of Study 2 samples were correctly classified with sensitivity 63.6%, specificity 72.7% when two latent components were used to predict cancer status with

adjustment for age and gender (Table S4A). However, the PLS analysis without age and sex adjustment yielded better separation of the groups for Study 2 samples (Figure 2D). Using two latent components meant 81.8% were correctly classified with sensitivity 72.7%, specificity 90.9%, suggesting an influence of covariants (Table S4B).

2.3. Detection of Unknown Metabolites

Some of the metabolites that we found to be differentially expressed were unknown compounds (Tables S1 and S2). We have intentionally used an untargeted metabolomic screen to detect novel metabolites that might be involved in the pathogenesis of NSLC and thereby detected “unknown” compounds in our metabolomic analysis. Most of these unknown compounds have been previously observed in other samples from different species, mammalian (human, mouse, rat), plant, bacterial and others which have been carefully tabulated in the BinBase database [25]. Searching our BinBase database and comparing the MS spectra of the unknowns with known compounds with similar electron ionization fragment spectra and similar retention times can help identify several interesting compounds that the unknown may be linked or is related to, based on the similarity of the spectra to known spectral fragmentation (Figure 3). For example, based on its fragmentation scan, the BinBase unknown compound #200595 has evidence of being an amino-compound, #200595 shows substructure patterns of carbohydrates and a retention index close to glucoheptose, #220177 can also be matched to carbohydrate substructures, in addition to showing a characteristic fragment m/z 144 typical for amines, and #223597 is a very high boiling compound with fragments found in sterols. Once additional cohort studies validate the importance of such unidentified metabolites, accurate mass GC-QTOF MS data can be acquired to obtain elemental formulas and matching structures from database queries [26,27].

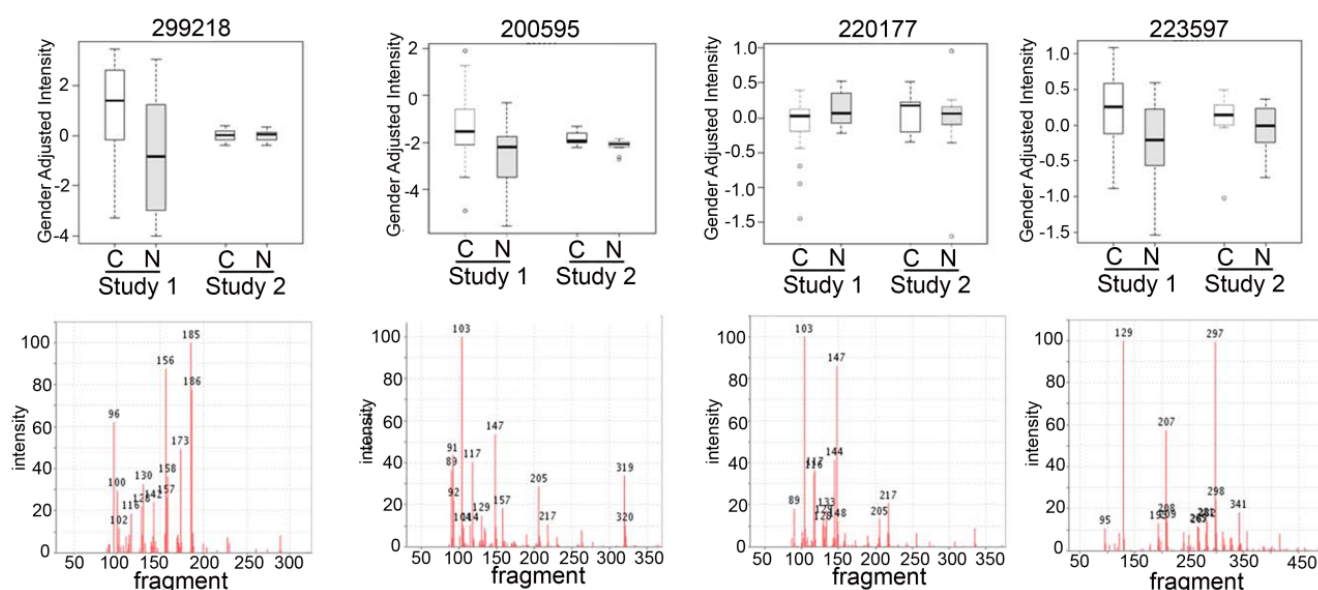


Figure 3. Box plots of top unknown compounds with electron ionization mass spectra comparing the two studies. Box-whisker plots (top panels) of the top unknown candidates from each study (Study 1 and Study 2) with the electron ionization MS spectra (lower panels) of the compound to show the mass fragmentation of the compounds to help with the identification of the compound.

2.4. Discussion

2.4.1. Systemic Metabolic Changes in Blood from Lung Cancer Patients

The most significant metabolite changes were representative of changes in amino acid, energy, fatty acid and lipid metabolism. Amino acid differences have been reported in NSCLC and other cancers in plasma [10,12,16,17,28] with Maeda *et al.* reporting that concentrations of Ser, Pro, Gly, Ala, Met, Ile, Leu, Tyr, Phe, Orn and Lys were higher and histidine was lower in NSCLC compared with controls [17], while Cascino *et al.* reported increases in Trp, Glu and Orn (Arg) in lung and breast cancers in blood [29]. Other amino acid changes we detected were consistent with a recent report by Miyagi *et al.* [30] who identified decreases in histidine, glutamine and threonine in early and late lung cancer in plasma samples from five cancers (lung, gastric, colorectal, breast and prostate). Miyagi *et al.* also reported that histidine was decreased in all but prostate cancer. Tryptophan was decreased in all five cancers in Miyagi's study [30], which we also identified as decreasing in our two studies. A general difference between the two studies is that Miyagi's study focused on the analysis of early stage cancers (Stage I–III) whereas Rossi Fanelli's studies analysed cancer anorexia in very late stage cancers [31,32].

Other studies in lung cancer in blood plasma have been conducted by NMR. Rocha *et al.* [33] detected metabolic changes related to glycolysis, glutaminolysis and gluconeogenesis with suppressed Krebs cycle and reduced lipid catabolism as a metabolic signature for lung cancer in 85 lung cancer patients and 78 healthy controls using NMR analysis. The average age for the lung cancer patients was 63 yr (30–85 yr) *versus* non-diseased control with an average age of 41 yr (22–60 yr) showing an age difference between the cases and controls. Histopathology of the cases ranged from 43% (37 cases) of adenocarcinoma with the rest a combination of epidemoid carcinoma, carcinoid, large cell and small cell carcinoma, with mostly early stage (81% Stage I and Stage II), and a few late stage (8% Stage III). Our Study 1 focused only on NSCLC adenocarcinoma with better matched controls for age and smoking history and all late stage disease. Hence, the metabolomic biomarkers we describe may be more associated with late stage, metastatic disease. We understand that the reason most experimental biomarkers fail in clinical validation studies for early detection is that the biomarkers were discovered in late stage cancer cases, and are unsuitable as biomarkers for early stage disease [34,35]. We still need to conduct studies on early stage lung cancer (Stage I–II) matched with suitable gender, age and smoking history controls to better characterize biomarkers of early lung cancer. Comparing the results from plasma and serum blood samples has already been shown to be marked similar for the two biofluids by Wedge *et al.* for small cell lung cancer (SCLC) [36]. Also, our studies complement those of the SIRM and NMR tissue and blood studies in lung cancer [17,22,37]. These studies require the infusion of C13 stable isotopes into the patient before analysis and are not practical for the clinical laboratory. The results from two studies add to the growing body of evidence that metabolomic changes detectable in blood and tissue could be used to detect and diagnose lung cancer.

2.4.2. Pathway Analysis and Overall Metabolic Effect on Blood Metabolites

While diagnosis of lung cancer phenotypes is clinically important, a differential analysis of plasma metabolic changes between lung cancer patients and matched controls should also reflect known mechanisms in cancer biology or lead to new hypotheses. We have therefore used less stringent thresholds

($p < 0.1$) to increase the number of metabolites that are potentially metabolically connected and visualized all metabolites at $p < 0.1$ using a combined biochemical and chemical network graph, MetaMapp (Figure 4). This graph clusters metabolites based on biochemical reactant pairs in the KEGG RPAIR Database [38,39] in addition to Tanimoto chemical structure similarity for identified metabolites that lack enzymatic information [40]. This type of analysis gives us a better perspective on how metabolic changes detected in blood samples might be due to the overall systemic effect of tumour growth and helps us to identify the potential biochemical links between the metabolic changes in blood from lung cancer (Figure 4). The use of MetaMAPP graphs enable understanding which metabolic modules are more affected by a disease or a treatment than others [40]. For example, few hydroxyl acids were differentially regulated, which means that trichloroacetic (TCA) cycle metabolites in blood plasma did not reflect the disease status, whereas plasma lactic acid was significantly increased in ACD NSLC patients. A range of both proteinogenic (tryptophan, lysine, histidine, valine) and non-proteinogenic amino acids (N-methylalanine, trans-4-hydroxyproline, cis-4-hydroxyproline) were found at lower levels in cancer patients, reflecting the increased use of carbon skeletons of amino acids in tumor cells. Reduced levels of amino acids have also been observed in cancer cachexia showing greater protein turnover and metabolism in advanced stage disease [41].

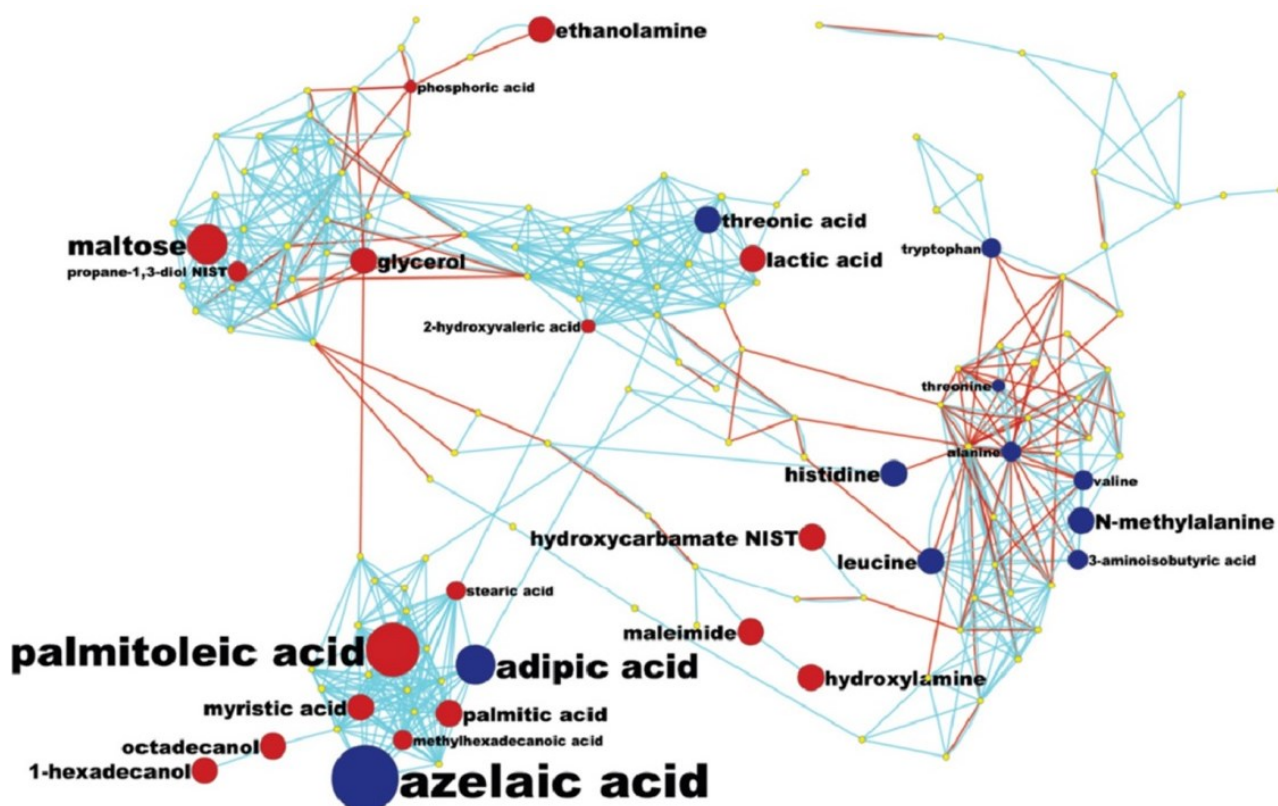


Figure 4. MetaMapp mapping of metabolomic analysis of lung cancer blood samples: a MetaMapp clustering metabolites based on biochemical reactant pairs in the KEGG RPAIR Database in addition to Tanimoto chemical structure similarity for identified metabolites that lack enzymatic information.

Interestingly, the most prevalent changes were seen in fatty acid/lipid biosynthesis (Figure 4). Several fatty acids were found to be up-regulated, including the important building blocks palmitate, stearate

and palmitoate, but also precursors for complex lipids such as glycerol and ethanolamine. The general increase in lipid biosynthesis is well established in tumor metabolism but has not been linked to plasma levels before. Conversely, peroxidative lipid breakdown products such as the dicarboxylic acids adipate and azelaic acid appear to be lowered in lung cancer patients, which may be linked to signalling functions or may indicate a lowered use of oxidative pathways in lipid metabolism in tumors. All of these possibilities are presently speculative and the actual mechanisms would need to be confirmed in more extensive studies.

2.4.3. Metabolomic Biomarker Potential for Lung Cancer Detection-Clinical Use

PLS and LDA were used on each study to determine if cancer and control samples could be separated and predicted based on metabolic profiles with data adjusted for sex and age or not. Using leave-one-out cross validation (LOOCV) of the results, which were regressed for sex and age, identified two latent components in Study 1 that were found to have sensitivity and specificity of 66.7% and 60% respectively and in Study 2 with a sensitivity of 63.6% and specificity of 72.7% (Tables S3A,B and S4A,B). We conclude from these results that our studies were not sufficiently robust to enable satisfactory biomarker identification. Our results do show similar results between cases and controls from two groups of samples. These results provide additional evidence for the use of metabolomic analysis to identify lung cancer biomarkers and support what others have already reported. Our results also add to our growing knowledge of the systems biology of lung cancer through our pathway analysis. However, considerably better performance characteristics than what we obtained here will be needed for clinical use. High specificity of biomarkers (>90% and 95%) will be necessary to adequately reduce false positive rates in lung cancer. It might be more feasible to use metabolomic biomarkers in the clinic to help monitor treatment and for recurrence. Again it will be important to have good sensitivity for this clinical application. We would need performance characteristics more similar to what is now being measured with current miRNA studies [42]. We are already conducting better designed studies with larger sample sizes with the goal of identifying suitable lung cancer biomarkers using metabolomic analysis. It may also be necessary to have a combination of different biomarkers or even a metabolomic profile to provide sufficient performance characteristics for clinical utility. The next hurdle would then be how best to adapt such an analysis to the clinical laboratory.

3. Experimental Section

3.1. Patient Samples

Pre-existing patient blood samples were acquired from the biorepositories of two institutions (FHCRC and UCDMC), all collected and stored at -80°C before use in these metabolomics studies. Samples were received with all identifiers removed. Each institution supplied an equal number of cases and controls (20 cases and 20 controls from FHCRC and 11 cases and 11 controls from UCDMC). Samples acquired from FHCRC were referred to as Study 1 (or FHCRC) and samples acquired from UCDMC were referred to as Study 2 (or UCDMC). All samples were collected with informed consent and following IRB protocols approved by each Institution's Institutional Review Board, intended for use only for research purposes. Collection and storage of clinical samples were conducted in accordance

with the Declaration of Helsinki with protocols approved by each institution (Gandara IRB protocol 255991-3). Blood samples (plasma) in Study 1 (FHCRC study) were taken from newly diagnosed lung cancer patients with NSCLC adenocarcinoma (mostly late stage) and were frequency matched with for age, gender and for general smoking history (current and former smokers) (Table 1A). Controls for this study were blood samples collected from individuals who were cancer free and with no history of cancer and who were current or former smokers. Blood (plasma) was collected using EDTA tubes (fasting conditions unknown), processed into plasma using approved protocols and stored at -80°C .

The second set of blood samples (Study 2) came from patients diagnosed with lung cancers (different types) (11 cases) that were frequency matched (age and gender) with samples from individuals without cancer and with no history of cancer (11 controls). These samples were obtained from the UC Davis Cancer Center Biorepository (CCB) and the UC Davis Clinical laboratory at the UC Davis Medical Center (UCDMC)(Table 1B). Smoking history and treatment status were known for most of the cases in Study 2, but not known for some of the control group. Fasting status of patients and controls were unknown. All blood samples (plasma and serum) were prepared using standard clinical SOPs specified at each institution and stored at -80°C until use. De-identified samples from each study were blinded and then subjected to metabolomic analysis as previously described [43–45] or described in the following methods section.

3.2. Non-Targeted Metabolomics Analysis by ALEX-CIS-GC/TOF MS

Samples were stored at -80°C prior to analysis. Samples were thawed and 30 μL of each sample was extracted and derivatized, and metabolite levels were quantified by gas chromatography time-of-flight (GC-TOF) mass spectrometry as previously described (38). Briefly, a 30 μL sample was extracted with 1 mL of degassed acetonitrile:isopropanol:water (3:3:2) at 20°C , centrifuged, the supernatant removed and solvents evaporated to dryness under reduced pressure. To remove membrane lipids and triglycerides, dried samples were reconstituted with acetonitrile/water (1:1), decanted and taken to dryness under reduced pressure. Internal standards, C8–C30 fatty acid methyl esters (FAMES), were added to samples and derivatized with methoxyamine hydrochloride in pyridine and subsequently by MSTFA (Sigma-Aldrich) for trimethylsilylation of acidic protons and analysed by GC-TOF mass spectrometry.

An Agilent 6890 gas chromatograph (Santa Clara, CA) was used with a 30 m, 0.25 mm i.d. Rtx5Sil-MS column with 0.25 μm 5% diphenyldimethylsiloxane film. A Gerstel MPS2 automatic liner exchange system was used to inject 0.5 μL of sample at 50°C (ramped to 250°C) with 25 s splitless injection time. An Agilent 6890 gas chromatograph (Santa Clara, CA, USA) was used with a 30 m, 0.25 mm i.d. Rtx5Sil-MS column with 0.25 μm 5% diphenyldimethylsiloxane film. An additional 10 m integrated guard column was used (Restek, Bellefonte, PA, USA). Chromatography was performed at a constant flow of 1 mL/min, ramping the oven temperature from 50°C to 330°C with a 22 min total run time. Mass spectrometry was conducted by a Leco Pegasus IV time of flight mass spectrometer with a 280°C transfer line temperature, electron ionization at -70 V and an ion source temperature of 250°C . Mass spectra were acquired from m/z 85–500 at 20 spectra s $^{-1}$ and 1750 V detector voltage. All samples were analyzed in one batch, throughout which data quality and instrument performance were monitored using quality control and reference plasma samples (National Institute of Standards and Technology, NIST).

3.3. Raw Data Processing and Chemometrics

Acquired spectra were further processed using the BinBase database [25,46]. Raw data files from the GC-TOF mass spectrometer were processed using proprietary software provided by the instrument company (Leco, ChromaTOF software (v. 2.32)) for peak finding and mass spectral deconvolution. Result files were exported and filtered for consistency using the UC Davis Metabolomics BinBase database. All metabolite spectra in BinBase [25] were matched against the Fiehnlib mass spectral library of 1200 authentic metabolite spectra using retention index and mass spectrum information in addition to the NIST05 commercial library. All output files were stored in the miniX study design database system. Identified metabolites were named by biochemical names, KEGG, referenced to PubChem as authoritative NIH/NCBI database and InChI hash keys encoding the chemical structures (Table S3). Further details of the BinBase algorithm and spectral libraries are given in Scholz and Fiehn [46] and Kind *et al.* [25].

3.4. Differential Analysis and Partial Least Squares Analysis

Prior to statistical analyses, metabolite intensities were mean normalized and log₂ transformed to meet underlying assumptions of normality with a constant variance and to reduce the dominant effect of extreme values. Differential analyses were conducted of each study (Study 1 and Study 2) to identify individual metabolites that differed between cancer and control subjects. For each study, a difference in mean intensity of each metabolite between cancer and control subjects was evaluated with an Analysis of Covariance (ANCOVA) which included gender and age as covariates. The significance of difference in each metabolite's intensity by cancer status was determined based on partial-F statistics using a parametric null distribution for each study. False discovery rates (FDRs) were calculated to account for multiple testing (Table S2).

Partial least squares (PLS) regression and linear discriminant analysis (LDA) [24,47] were conducted on the metabolomic data to determine if cancer and control patients could be separated and thereby predict the sample classification based on global metabolomic profiles. PLS regression was performed to reduce the intensity measures of the total number of 437 peaks to a smaller number of latent components that explained most of the variation in the data. Leave-one-out cross validation (LOOCV) was used to determine the optimal number of latent components to use for the LDA for each study. Through LOOCV, the class membership of each excluded subject (cancer or control) was predicted with LDA using 1 through 10 latent components identified with the other subjects. For each number of latent components evaluated, the misclassification rate was calculated. The number of latent components yielding the lowest misclassification rate was considered optimal.

Each study was analysed separately. We conducted the PLS-LDA for each study in two ways. Because gender and age could be confounding factors, in the first analysis we adjusted for these factors by using the residuals from a linear regression of metabolite intensities *versus* gender and age in the PLS-LDA. In the second analysis, metabolite intensities were not adjusted for age and gender. For all analyses, the misclassification rate, sensitivity and specificity of the PLS-LDA for 1 through 10 latent components were calculated through leave-one-out cross validation (Tables S3A,B and S4A,B). In

addition, R², Q², the percent variation in the metabolite intensities explained by the PLS regression were calculated. PLS and LDA analyses were conducted with the R package *pls*genomics [48].

3.5. MetaMapp Mapping of Identified Compounds

Molfile encoded chemical structures were retrieved from the PubChem database using PubChem identifiers of the identified metabolites. The structures were subjected to pairwise Tanimoto chemical similarity coefficient calculations using an online structure clustering tool available at PubChem website. For this analysis we relaxed the stringency of the statistical analysis of metabolites and included those with $p < 0.1$ to increase the number of metabolites that are potentially metabolically connected. This graph clusters metabolites based on biochemical reactant pairs in the KEGG RPAIR Database [38] in addition to Tanimoto chemical structure similarity for identified metabolites that lack enzymatic information [40]. The resulting similarity matrix was downloaded and used as an input for MetaMapp mapping software [40] at www.metamapp.fiehnlab.ucdavis.edu. A network graph of one-reaction steps for the metabolites that were annotated with KEGG identifiers was calculated using MetaMapp software. Generated network graphs were imported into Cytoscape and were visualized using organic layout algorithm. Results of ANOVA statistics were converted into a node-attributed file using an online tool available on the MetaMapp website. Directions of differential alterations were visualized as node color and fold-changes were visualized as node size.

4. Conclusions

We analysed blood samples from lung cancer cases and matched controls to determine whether metabolomic analysis of blood by GC-MS has the potential to be used to detect the presence of lung cancer. Similar metabolic differences were detected in two pilot studies. Generally, energy or carbohydrate metabolites (maltose, glycerol, lactic acid) increased, amino acids (tryptophan, lysine, histidine decreased and certain fatty acids (palmitic acid) increased in both studies.

What is valuable about our studies is the overall finding that metabolomic analysis has the potential to distinguish blood samples of patients with lung cancer from those without cancer. This type of analysis has potential for clinical use to aid diagnosis, especially if it can help identify different types of lung cancer (NSCLC from SCLC or adenocarcinoma from squamous), which could impact the type of treatment. Furthermore if metabolomic biomarkers and a systems biology approach can be proven to have high specificity, this type of analysis/assay could possibly help reduce the high false positive rate for low-dose CT screening. Clinical use would require adaptation to the clinical laboratory, which we foresee would not be difficult since there are already mass spectrometers in clinical chemistry being used to analyse specific chemicals and drugs. Target assays for specific metabolites could be developed for the clinical laboratory and used for blood testing.

These findings support the idea that metabolic changes measurable in blood samples are indications of early systemic metabolic consequences of lung cancer, which could potentially be useful in a clinical setting to enhance present diagnostic methods. In addition to aiding diagnosis of lung cancer, metabolomic and other types of biomarkers analysis (proteomics, glycomics, miRNA) could accompany present imaging methods such as low-dose CT scan screening to reduce its high false positive rate for malignant lung cancer if early metabolic changes related to progression of cancer can be detected in

blood samples. Further investigations into the use of metabolomic analyses of blood samples in much larger clinical studies are needed in order to adequately assess the clinical value of this type of analysis.

Acknowledgments

Funding support - DOD LCRP W81XWH-10-1-0635 to D.R. Gandara PI, R21 CA135240-01A1 to S. Miyamoto. This study was supported in part by an award (to Miyamoto) from LUNgevity Foundation and the Thomas G. Labrecque Foundation; EU FP7 Health-2007# 200327 and National Institutes of Health, RC2 GM092729 to O. Fiehn. We acknowledge the help of Sevini Shahbaz in the Fiehn laboratory with metabolomic analysis of the samples and the efforts of Stephanie Soares, Irmgard Feldman and Regina Gandour-Edwards of the UC Davis Cancer Biorepository for their assistance with acquiring the UCDCMC samples.

Author Contributions

Suzanne Miyamoto contributed to the study design, procured samples and assisted with the process of the samples in the metabolomic analysis (Study 2), contributed to the data analysis and managed the organization of the research project; was largely responsible for the writing of the manuscript.

Sandra L. Taylor performed statistical analysis of the data results.

Dinesh K. Barupal prepared the MetaMAPP figure (technical expertise).

Ayumu Taguchi contributed to interpretation of data results and the preparation of the manuscript.

Gert Wohlgemuth programmed software to detect metabolites and report spectra for unidentified compounds.

Ken Y. Yoneda and David R. Gandara provided important clinical expertise to this project and helped with the collection of clinical samples.

William R. Wikoff provided expertise with metabolomic analysis and contributed to the preparation of the manuscript.

Samir M. Hanash provided samples (Study 1), clinical and experimental expertise, and preparation of the manuscript.

Kyoungmi Kim contributed to the study design and performed statistical analysis and contributed to the preparation of the manuscript.

Oliver Fiehn provided conceptual and technical guidance, contributed to the study design, provided the laboratory equipment and expertise to conduct metabolomic analysis, contributed to the interpretation of the data results, and contributed to the preparation of the manuscript.

Supplementary Materials

Supplementary materials can be accessed at: www.mdpi.com/2218-1989/5/2/192/s1

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Siegel, R.; Ma, J.; Zou, Z.; Jemal, A. Cancer statistics, 2014. *CA Cancer J. Clin.* **2014**, *64*, 9–29.
2. Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *CA Cancer J. Clin.* **2011**, *61*, 69–90.
3. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* **2008**, *83*, 584–594.
4. National Lung Screening Trial Research, T.; Aberle, D.R.; Berg, C.D.; Black, W.C.; Church, T.R.; Fagerstrom, R.M.; Galen, B.; Gareen, I.F.; Gatsonis, C.; Goldin, J.; *et al.* The national lung screening trial: Overview and study design. *Radiology* **2011**, *258*, 243–253.
5. Goulart, B.H.; Bensink, M.E.; Mummy, D.G.; Ramsey, S.D. Lung cancer screening with low-dose computed tomography: Costs, national expenditures, and cost-effectiveness. *J. Natl. Compr. Canc. Netw.* **2012**, *10*, 267–275.
6. Louie, A.V.; Rodrigues, G.; Olsthoorn, J.; Palma, D.; Yu, E.; Yaremko, B.; Ahmad, B.; Aivas, I.; Gaede, S. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother. Oncol.: J. Eur. Soc. Ther. Radiol. Oncol.* **2010**, *95*, 166–171.
7. Bigbee, W.L.; Gopalakrishnan, V.; Weissfeld, J.L.; Wilson, D.O.; Dacic, S.; Lokshin, A.E.; Siegfried, J.M. A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by ct screening. *J. Thorac. Oncol.* **2012**, *7*, 698–708.
8. Tsay, J.C.; DeCotiis, C.; Greenberg, A.K.; Rom, W.N., Current readings: Blood-based biomarkers for lung cancer. *Semin. Thorac. Cardiovasc. Surg.* **2013**, *25*, 328–334.
9. Deberardinis, R.J.; Sayed, N.; Ditsworth, D.; Thompson, C.B. Brick by brick: Metabolism and tumor cell growth. *Curr. Opin. Genet. Dev.* **2008**, *18*, 54–61.
10. Sreekumar, A.; Poisson, L.M.; Rajendiran, T.M.; Khan, A.P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R.J.; Li, Y.; *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009**, *457*, 910–914.
11. Denkert, C.; Budczies, J.; Kind, T.; Weichert, W.; Tablack, P.; Sehouli, J.; Niesporek, S.; Konsgen, D.; Dietel, M.; Fiehn, O. Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res.* **2006**, *66*, 10795–10804.
12. Urayama, S.; Zou, W.; Brooks, K.; Tolstikov, V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 613–620.
13. Fan, L.; Zhang, W.; Yin, M.; Zhang, T.; Wu, X.; Zhang, H.; Sun, M.; Li, Z.; Hou, Y.; Zhou, X.; *et al.* Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a uplc/qtof/ms platform. *Acta Oncol.* **2012**, *51*, 473–479.
14. Slupsky, C.M.; Steed, H.; Wells, T.H.; Dabbs, K.; Schepansky, A.; Capstick, V.; Faught, W.; Sawyer, M.B. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer Res.* **2010**, *16*, 5835–5841.
15. Woo, H.M.; Kim, K.M.; Choi, M.H.; Jung, B.H.; Lee, J.; Kong, G.; Nam, S.J.; Kim, S.; Bai, S.W.; Chung, B.C. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. *Clin. Chim. Acta* **2009**, *400*, 63–69.

16. Yu, L.; Aa, J.; Xu, J.; Sun, M.; Qian, S.; Cheng, L.; Yang, S.; Shi, R. Metabolomic phenotype of gastric cancer and precancerous stages based on gas chromatography time-of-flight mass spectrometry. *J. Gastroenterol. Hepatol.* **2011**, *26*, 1290–1297.
17. Maeda, J.; Higashiyama, M.; Imaizumi, A.; Nakayama, T.; Yamamoto, H.; Daimon, T.; Yamakado, M.; Imamura, F.; Kodama, K. Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: A case control study. *BMC Cancer* **2010**, *10*, 690.
18. Amon, L.M.; Pitteri, S.J.; Li, C.I.; McIntosh, M.; Ladd, J.J.; Disis, M.; Porter, P.; Wong, C.H.; Zhang, Q.; Lampe, P.; *et al.* Concordant release of glycolysis proteins into the plasma preceding a diagnosis of er+ breast cancer. *Cancer Res.* **2012**, *72*, 1935–1942.
19. Buckendahl, A.C.; Budczies, J.; Fiehn, O.; Darb-Esfahani, S.; Kind, T.; Noske, A.; Weichert, W.; Sehouli, J.; Braicu, E.; Dietel, M.; *et al.* Prognostic impact of amp-activated protein kinase expression in ovarian carcinoma: Correlation of protein expression and gc/tof-ms-based metabolomics. *Oncol. Rep.* **2011**, *25*, 1005–1012.
20. Suhre, K.; Shin, S.Y.; Petersen, A.K.; Mohny, R.P.; Meredith, D.; Wagele, B.; Altmaier, E.; CardioGram; Deloukas, P.; Erdmann, J.; *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **2011**, *477*, 54–60.
21. Serkova, N.J.; Glunde, K. Metabolomics of cancer. *Methods Mol. Biol.* **2009**, *520*, 273–295.
22. Lane, A.N.; Fan, T.W.; Bousamra, M., 2nd; Higashi, R.M.; Yan, J.; Miller, D.M. Stable isotope-resolved metabolomics (SIRM) in cancer research with clinical application to nonsmall cell lung cancer. *OMICS* **2011**, *15*, 173–182.
23. Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate Analysis*; Academic Press: London, UK; New York, NY, USA, 1979; p 15.
24. Wold, S. Exponentially weighted moving principal components-analysis and projections to latent structures. *Chemometr. Intell. Lab. Syst.* **1994**, *23*, 149–161.
25. Kind, T.; Wohlgemuth, G.; Lee do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. Fiehnlib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **2009**, *81*, 10038–10048.
26. Kumari, S.; Stevens, D.; Kind, T.; Denkert, C.; Fiehn, O. Applying in-silico retention index and mass spectra matching for identification of unknown metabolites in accurate mass gc-tof mass spectrometry. *Anal. Chem.* **2011**, *83*, 5895–5902.
27. Kind, T.; Fiehn, O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinf.* **2007**, *8*, 105.
28. Wen, T.; Gao, L.; Wen, Z.; Wu, C.; Tan, C.S.; Toh, W.Z.; Ong, C.N. Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. *Mol. Biosyst.* **2013**, *9*, 2370–2378.
29. Cascino, A.; Muscaritoli, M.; Cangiano, C.; Conversano, L.; Laviano, A.; Ariemma, S.; Meguid, M.M.; Rossi Fanelli, F. Plasma amino acid imbalance in patients with lung and breast cancer. *Anticancer Res.* **1995**, *15*, 507–510.
30. Miyagi, Y.; Higashiyama, M.; Gochi, A.; Akaike, M.; Ishikawa, T.; Miura, T.; Saruki, N.; Bando, E.; Kimura, H.; Imamura, F.; *et al.* Plasma free amino acid profiling of five types of cancer patients and its application for early detection. *PLoS ONE* **2011**, *6*, e24143.

31. Rossi Fanelli, F.; Cangiano, C.; Ceci, F.; Cellerino, R.; Franchi, F.; Menichetti, E.T.; Muscaritoli, M.; Cascino, A. Plasma tryptophan and anorexia in human cancer. *Eur. J. Cancer Clin. Oncol.* **1986**, *22*, 89–95.
32. Rossi Fanelli, F.; Cangiano, C.; Muscaritoli, M.; Conversano, L.; Torelli, G.F.; Cascino, A. Tumor-induced changes in host metabolism: A possible marker of neoplastic disease. *Nutrition* **1995**, *11*, 595–600.
33. Rocha, C.M.; Carrola, J.; Barros, A.S.; Gil, A.M.; Goodfellow, B.J.; Carreira, I.M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; *et al.* Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. *J. Proteome Res.* **2011**, *10*, 4314–4324.
34. Zhu, C.S.; Pinsky, P.F.; Cramer, D.W.; Ransohoff, D.F.; Hartge, P.; Pfeiffer, R.M.; Urban, N.; Mor, G.; Bast, R.C., Jr.; Moore, L.E.; *et al.* A framework for evaluating biomarkers for early detection: Validation of biomarker panels for ovarian cancer. *Cancer Prev. Res.* **2011**, *4*, 375–383.
35. Cramer, D.W.; Bast, R.C., Jr.; Berg, C.D.; Diamandis, E.P.; Godwin, A.K.; Hartge, P.; Lokshin, A.E.; Lu, K.H.; McIntosh, M.W.; Mor, G.; *et al.* Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev. Res.* **2011**, *4*, 365–374.
36. Wedge, D.C.; Allwood, J.W.; Dunn, W.; Vaughan, A.A.; Simpson, K.; Brown, M.; Priest, L.; Blackhall, F.H.; Whetton, A.D.; Dive, C.; *et al.* Is serum or plasma more appropriate for intersubject comparisons in metabolomic studies? An assessment in patients with small-cell lung cancer. *Anal. Chem.* **2011**, *83*, 6689–6697.
37. Fan, T.W.; Lane, A.N.; Higashi, R.M.; Farag, M.A.; Gao, H.; Bousamra, M.; Miller, D.M. Altered regulation of metabolic pathways in human lung cancer discerned by ¹³C stable isotope-resolved metabolomics (SIRM). *Mol. Cancer* **2009**, *8*, 41.
38. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114.
39. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360.
40. Barupal, D.K.; Haldiya, P.K.; Wohlgemuth, G.; Kind, T.; Kothari, S.L.; Pinkerton, K.E.; Fiehn, O. Metamapp: Mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinf.* **2012**, *13*, 99.
41. Pisters, P.W.; Pearlstone, D.B. Protein and amino acid metabolism in cancer cachexia: Investigative techniques and therapeutic interventions. *Crit. Rev. Clin. Lab Sci.* **1993**, *30*, 223–272.
42. Sozzi, G.; Boeri, M.; Rossi, M.; Verri, C.; Suatoni, P.; Bravi, F.; Roz, L.; Conte, D.; Grassi, M.; Sverzellati, N.; *et al.* Clinical utility of a plasma-based mirna signature classifier within computed tomography lung cancer screening: A correlative mild trial study. *J. Clin. Oncol.* **2014**, *32*, 768–773.
43. Weckwerth, W.; Wenzel, K.; Fiehn, O. Process for the integrated extraction, identification and quantification of metabolites, proteins and rna to reveal their co-regulation in biochemical networks. *Proteomics* **2004**, *4*, 78–83.
44. Fiehn, O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Anal. Chem.* **2008**, *27*, 261–269.
45. Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R.H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **2007**, *363*, 185–195.

46. Scholz, M.; Fiehn, O. Setupx—A public study design database for metabolomic projects. *Pac. Symp. Biocomput.* **2007**, *363*, 169–180.
47. Boulesteix, A.L.; Strimmer, K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform* **2007**, *8*, 32–44.
48. R Development Core Team. The R Stats Package. Available online: <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/00Index.html> (accessed on 31 March 2015).

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).